DOCUMENT RESUME

ED 110 462 TM 004 696

AUTHOR Willson, Victor L.
TITLE A Two-Factor ANOVA Model for Time-Series Experiments.
PUB DATE [Apr 75]
NOTE 17p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, D.C., March 30-April 3, 1975)

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS *Analysis of Variance; *Classification; Correlation; Intervention; *Mathematical Models; Models; Observation; *Research Methodology; Statistical Analysis; Test Results; Time

ABSTRACT

A theoretical model is developed for analysis of intervention effects in multiple autocorrelated time-series which may be classified or blocked two ways. The intervention effect is estimated independently for each series. The estimates are then entered as observations in two-factor analysis of variance with one observation per cell, utilizing a pooled variance error estimate for the mean square error. The model was tested empirically through the use of Monte Carlo techniques, and results are presented in tabular form for Type I and Type II error levels. An example is presented with data taken from a Kansas City Police study utilizing the two-factor layout. (Author)

**********
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
**********

ED110462

# A Two-Factor ANOVA Model for Time-Series Experiments

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Victor L. Willson
The University of Minnesota

Presented at the Annual Meeting of the American Educational Research Association, Washington, D.C., April 2, 1975.

TM 004 696

# A Two-Factor ANOVA Model for Time-Series Experiments[1]

The interrupted time-series experiment was discussed by Campbell and Stanley (1963) as a quasi-experimental alternative to experimental designs in certain behaviorial science settings. Box and Tiao (1965) developed a statistical model for estimating changes in the level of a class of time-series processes which corresponded to the Campbell-Stanley conceptual model. The concepts have been generalized and refined by several researchers in the past decade and many results have been presented in Glass, Willson, and Gottman (1975). Discussion of the time-series experiment for small sample size (< 500) has been limited to models of complexity no greater than the one-way classification in which a set of time-series is examined and contrasts over the parameters estimated constructed. This paper details the extension of the theory of time-series experiments to two-way layouts, and by extrapolation, to the multi-way classification schemes.

## Models for Time-Series

Box and Jenkins (1970) have outlined a general procedure for representing time series as autoregressive-integrated moving average (ARIMA) processes, working in the "time" domain, as opposed to the "frequency" domain of spectral analysis. ARIMA models are linear and have deterministic and stochastic components. Their salient feature is the structural representation of cross-time correlations of observations with themselves. It has been this feature which has provided difficulty in analysis of variance, as those

[1] The author wishes to thank Dr. Tony Pate, Kansas City Police Foundation, and Mr. Duane Dieckman, Midwest Research Institute, for generating the problem and providing the data used in the example.

familiar with repeated measures analysis can attest. Ordinary parametric statistical analysis is inappropriate because the distributions of statistics of interest are not as expected (cf. Scheffe, 1959, for a discussion of serial correlation, a simple example of time-series correlation).

The Box-Jenkins model ARIMA (p,d,q) is a general representation

$$\phi_p(B)\nabla^d z_t = \theta_q(B)a_t , \qquad (1)$$

in which $\phi_p$ and $\theta_q$ are exponential polynomials in B of order p and q, $z_t$ is an observation of a time-series at time $t$ ; $\nabla$ is a difference operator such that $\nabla z_t = z_t - z_{t-1}$ ; B is a Shift operator in time $t$, e.g., $B(z_t) = z_{t-1}$; $a_t$ is a random variable distributed independently for each $t$ and with identical variance for all $t$; $L$ is the mean level of the series when first observed.[2]

This complicated model may be divided into two components; one prior to intervention and one posterior to intervention. An intervention effect, $\delta$, represents a change in experimental condition, and in its simplest form can be considered a deterministic change in the level $L$ of the series. Box and Tiao (1965) derived an unbiased estimate for the intervention effect with known statistical properties for one class of time series. Kepka (1972) generalized this derivation for the whole ARIMA class. The general procedure involves finding an appropriate transformation for the time series after the ARIMA model parameters $p$, $d$, and $q$ have been specified. The transformed series has the form of the general linear model and ordinary least-squares estimation is made. Assumption of normality of the random unknown variables $a$ allows testing of hypotheses since the statistics divided by their standard errors

[2] In general, these series do not have expectations in the first or second moments, and $L$ is just an origin point, not having the same property as a true mean.

then follow t-distributions.

In particular;

$$(\hat{\delta} - \delta) / \hat{\sigma}_{\hat{\delta}} \cdot \sqrt{(X'X)^{-1}_{ii}} \sim t_{N-p}, \quad (2)$$

where $(X'X)^{-1}_{ii}$ is the $i$-th diagonal element of the matrix $(X'X)^{-1}$ commonly associated with least squares estimation in the general linear model $Y = X\beta + e$ ; $p$ is the number of independent parameters estimated. See Glass, Willson, and Gottman (1975) for a derivation of the estimators.

Shumway (1970) has discussed the general problem of estimating regression and ANOVA models for time-series in the frequency domain of spectral analysis. The procedures he outlined seem adequate for the problems discussed in this paper, but the number of observations for each time series suggested for utility of the approach is over 500. Since such a number is found in few behavioral time lines, the time domain approach detailed below is offered as a "small" sample alternative ( $50 < n < 500$ for each time-series).

## A Two-Way Classification Design

Glass (1968) discussed a case in which several intervention effects were examined in post hoc multiple comparisons using Scheffe's procedure (1959). The tests were based on the fact that the distributions of the statistics ($\hat{\delta}_i$) were known for each case. An analogous situation exists when a two-way classification is made. Each cell in an I x J layout represents a time-series with $n_{ij}$ observations. For each series an intervention has been made after $n_{1,ij}$ observations. Only a simple $\delta_{ij}$, change in level will be considered, but results are easily extended to cases where $\delta_{ij}$ is a more complicated effect (such as being a function of time).

A conceptually useful model is that taken from analysis of variance:

$$\delta_{ij} = \delta + \alpha_i + \beta_j + \alpha\beta_{ij} + e_{ij} \quad , \qquad (3)$$

in which $\delta$ is the general change in the population of time-series, and $\alpha_i$ , $\beta_j$, $\alpha\beta_{ij}$ are the contributions due to inclusion in the levels of the two classification factors, labeled A and B.

The population on which the time-series observations are made (or populations) must have logical similarities. The series might be observations made on humans, crime statistics, inflation rates, etc. The classification scheme produces a causal framework in which the researcher is interested in testing effects of the classification variables or merely in accounting for variance in the intervention effects.

Each series is analyzed separately to obtain least-squares estimates for $\delta_{ij}$. This is analogous to estimating a mean for each cell in an ANOVA design for difference scores.

The difference scores here are of a special sort, however, and do not lend themselves to computing a total sum of squares. Information is available on the variance error of the $\delta_{ij}$ estimate, $\hat{\delta}_{ij}$, and the distribution of $\hat{\delta}_{ij}$ is known to be normal; this information allows estimation of all parameters in the model given by (3).

If the complexity of the problem were no more than that mentioned above, the paper would warrant little more than a note. Some major unique features of a time-series design must be examined, however. One is that for each time series analyzed, the series' ARIMA process must be assumed or identified. The estimation of $\delta_{ij}$ and $\sigma^2_{\tilde{\delta}_{ij}}$ both rely on the specification of the ARIMA parameters p, d, q; Once the order of process is known, the parameters of $\phi_p(B)$ and $\Theta_q(B)$ must be either given or estimated. Thus, several major sources of error are introduced.

The identification problem has received a significant amount of attention by Bartlett (1946), Box and Jenkins (1970), Durbin and Watson (1950), and many others. In general, the errors made in identification are reduced as the number of observations increases. One effect of misspecifying a process is to increase error variance of the estimated parameter greatly.

The consequences of incorrectly estimating the parameters of $\phi_p(B)$ and $\theta_q(B)$ are not well known. A doctoral student[3] at the University of Minnesota is currently examining the distributions of parameters estimates for sampling distributions of the $\phi$ and $\theta$ parameters.

A second class of problems involves the threats to internal and external validity of a two-way layout of time-series. No requirement has been made that series are observed at the same time, that interventions occur simultaneously, or even that intervals of time between observations are the same for each series.[4] A threat due to historical anomalies occurring in at least some intervention intervals is increased if observations occur in different real time intervals. That is, the chances of some unrelated factor producing a spurious change in at least one of the series is increased when the span of time for conducting the experiment increases.

The change in interval of observation from one cell to the next is not so bizarre as it may seem. The same statistic may be collected weekly for one unit, monthly for another. If the researcher has control over the observation schedule he would be advised to keep the interval identical for all cells.

[3]Gerald Martin, Department of Educational Administration, College of Education, University of Minnesota, Minneapolis.

[4]All writers assume that within a series observations are made at equal intervals; this is a restriction that should be examined, and results derived to remove it; why? missing data, unequal observation schedules, etc., are not uncommon.

If the interval is great for some cells, the intervention effect may be lost, especially if it is ephemeral[5] (dies out quickly), or the power of the tests is low because the number of observations is reduced.

Problems of reliability of observation may be involved when series are sequential, as in clinical psychology, when observations are made for a series of clients. Ratings may change as raters become more familiar with the process, fatigued, careless, etc. One alternative when such a problem is involved is to include <u>time</u> as a factor! A trend analysis on the $\hat{\sigma}^2_{\delta_{ij}}$ will indicate if this is a possibility, or the whole experiment may be considered a single time-series and a test for trend made.

This discussion does not exhaust the possible time-related problems of a two-way layout, but it does suggest that many of the problems will be avoided when the researcher is able to plan the experiment rather than rely on a post hoc experiment. When the latter case exists, he must consider the serious problems which may accompany it.

## Equal Cell Sizes

When all $N_{ij} = n$ and all $N_{1,ij} = n_1$ (See Class, Willson, and Gottman, 1975, about the problem of when to intervene in a series), the analysis of variance model is seen to be an $I \times J$ model with one "observation" per cell, $\hat{\delta}_{ij}$. The error term for each of the mean squares for effects is a pooled variance estimate,

$$\hat{\sigma}^2_e = \sum_{i=1}^{I} \sum_{j=1}^{J} \hat{\sigma}^2_{\delta_{ij}} / IJ \qquad (4)$$

[5] Even though the model in (3) assumes a permanent change in level of the series, analysis is almost identical for ephemeral interventions, and it may be impossible to tell such changes apart in practice.

The ANOVA Table appears as Table 1. The number of observations per cell merely weight the estimate $\hat{\delta}_{ij}$ for analysis purposes. For equal or proportional cell sizes the interpretations of statistics are straightforward. The F-ratios are all central-distributed under null hypotheses on the effects with the degrees of freedom given in Table 1. Assumptions are homogeneity of variance across cells (which may be tested), normality of errors, and independence of errors.[6]

The error in (3) is a function of the error in the original time-series, which was assumed to be random and uncorrelated with errors in any other series.

Processes have not been assumed identical across cells of the design That is, the parameters p, d, ,q, $\phi(B)$, or $\theta(B)$ need not be identical, since the parameter of interest, $\delta_{ij}$, is estimable regardless of the equivalence br nonequivalence of the other parameters. If the models change drastically from cell to cell, however, this may be an indication of logical inconsistency, perhaps associated with the original classification scheme. One would expect subjects and units to behave somewhat the same statistically across time series.

## Disproportionate Cell Sizes

The two-way (or higher) layout seems amenable to analysis when the numbers of observations taken for the cells are disproportionate. Since there is actually only one observation, $\hat{\delta}_{ij}$, per cell, the disproportionality is a characteristic of the information utilized to estimate $\delta_{ij}$. Thus, an unweighted means analysis seems useful.

[6] Non-normality seems to have little effect on estimating $\delta_{ij}$. A doctoral thesis by Judith George Doerann (1973) at University of Minnesota investigated violations of normality. Glass, Gullickson, and Nelson (1971) investigated nonhomogeneity of variance for estimating $\delta_{ij}$.

Table 1: Two-Way Classification ANOVA Table for Equal Cell Sizes, Three Models (Fixed, Random, and Mixed)

| Source | Sum of Squares | Degrees of Freedom | E (Mean Square) Model | | |
|---|---|---|---|---|---|
| | | | $A_1B$ Fixed | $A_1B$ Random | Mixed A fixed B random |
| A | $J\sum_i (\bar{\delta}_{i.} - \bar{\delta}_{..})^2$ | I-1 | $\sigma_e^2 + J\sigma_\alpha^2$ | $\sigma_e^2 + J\sigma_\alpha^2 + \sigma_{\alpha\beta}^2$ | $\sigma_e^2 + J\sigma_\alpha^2 + \sigma_{\alpha\beta}^2$ |
| B | $I\sum_j (\bar{\delta}_{.j} - \bar{\delta}_{..})^2$ | J-1 | $\sigma_e^2 + I\sigma_\beta^2$ | $\sigma_e^2 + I\sigma_\beta^2 + \sigma_{\alpha\beta}^2$ | $\sigma_e^2 + I\sigma_\beta^2$ |
| A x B | $\sum_i\sum_j (\delta_{ij} - \bar{\delta}_{i.} - \bar{\delta}_{.j} + \bar{\delta}_{..})^2$ | (I-1)(J-1) | $\sigma_e^2 + \sigma_{\alpha\beta}^2$ | $\sigma_e^2 + \sigma_{\alpha\beta}^2$ | $\sigma_e^2 + \sigma_{\alpha\beta}^2$ |
| error | $\sum_i\sum_j \hat{\sigma}_{\delta_{ij}}^2 / IJ$ | (N-p)IJ | $\sigma_e^2$ | $\sigma_e^2$ | $\sigma_e^2$ |

$$\bar{\delta}_{i.} = \sum_j \delta_{ij} / J$$

$$\bar{\delta}_{.j} = \sum_i \delta_{ij} / I$$

$$\bar{\delta}_{..} = \sum_i\sum_j \delta_{ij} / IJ$$

p = number of parameters estimated in the estimation of $\delta_{ij}$

The harmonic mean of cell size is

$$\bar{n}_h = \frac{IJ}{\sum_{i=1}^{I}\sum_{j=1}^{J}\left(\frac{1}{n_{ij}-p}\right)} \quad . \qquad (5)$$

Following Winer (1971) The sums of squares are

$$SS_A = J\sum_{i=1}^{I}(\bar{\delta}_{i.}-\bar{\delta}_{..})^2 \quad , \qquad (6)$$

$$SS_B = I\sum_{j=1}^{J}(\bar{\delta}_{.j}-\bar{\delta}_{..})^2 \quad , \qquad (7)$$

$$SS_{AXB} = \sum_{i=1}^{I}\sum_{j=1}^{J}(\delta_{ij}-\bar{\delta}_{i.}-\bar{\delta}_{.j}+\bar{\delta}_{..})^2 \quad , \qquad (8)$$

where the statistics are identical to those in Table 1. The error variance is a weighted pooled estimate,

$$\hat{\sigma}_e^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\nu_{ij}\hat{\sigma}^2_{\delta_{ij}}/IJ, \qquad (9)$$

where

$$\nu_{ij} = (n_{ij}-p)/\bar{n}_h$$

The degrees of freedom associated with error are

$$df_{error} = \sum_{i=1}^{I}\sum_{j=1}^{J}(n_{ij}-p) - IJ+1 \qquad (10)$$

The equal number of statistics (1) for each level of the experiment motivates this analysis. The $n_{ij}$ are merely the relative weight for error.

## Monte Carlo Study

As a rough check 1000 experiments were generated using Control Data Corporation's CYBER 75 machine at the University of Minnesota.[7] Each experiment consisted of a 2 x 3 layout with 60 observations per cell. An intervention effect three times the random error variance in magnitude was introduced after 30 time-series observations in each cell. No treatment effects were included, so that conditions were those of "no difference" null hypotheses. All series were IMA (1,1) processes with parameter $\theta_1 = .5$. Each series in each experiment was analyzed and estimates for known $\theta_1$ of $\delta_{ij}$ and $\sigma^2_{\delta_{ij}}$ obtained. These were statistics in the ANOVA for equal cell sizes presented above. F-statistics for treatment effects and interaction were computed, along with the variance error estimate.

Results are given in Table 2. The theoretical F-distributions and observed F-distributions are compared in terms of the cumulative proportions (probabilities) of the statistics. The correspondence is certainly not perfect, but is at least an indication of the correctness of the procedure.

[7] Random numbers were generated using Subroutine Normal at University of Minnesota.

Table 2: Results for a 2 x 3 Time Series Design Monte Carlo Experiment of F-distributions

| | Effect | | |
|---|---|---|---|
| | A | B | A x B |
| | df= 1,348 | df= 2,348 | df= 2,348 |
| Theoretical Proportion of Cumulative Distribution | Observed proportion | Observed Proportion | Observed Proportion |
| .05 | .04 | .05 | .04 |
| .10 | .12 | .10 | .09 |
| .25 | .32 | .22 | .25 |
| .50 | .62 | .47 | .48 |
| .75 | .88 | .73 | .71 |
| .90 | .96 | .87 | .85 |
| .95 | .98 | .93 | .91 |

Illustrative Example

As reported by Kelling et al (1974), The Kansas City Police Foundation funded a police deployment experiment in Kansas City between October 1972, and September 1973. Fifteen police "beats" were assigned to three treatments -- proactive, reactive, and control—so that each treatment had five levels. Beats were assigned so as to be non-contiguous within treatment and were matched on socioeconomic conditions. The reactive treatment eliminated routine patrol in favor of a response-only enforcement; proactive treatment consisted of intensified patrolling (two to three times the control level, which was one patrol car per beat).

Data were available by month for four years prior to the experiment, and after the "intervention" twelve months' data were collected, the time of the experimental conditions. Data on burglaries in the fifteen beats were used to present an example of the procedures detailed earlier and were seasonally adjusted prior to analysis.

The analysis of the data is presented in Table 3. Since beats is a random factor, the analysis model is of mixed type. The treatment effect is seen to be significant at $p < .10$. As noted by Glass, Willson, and Gottman (1975), one can use multiple comparisons techniques to examine specific contrasts of interest. Kelling et al (1974) used this procedure to compare treatment means averaged over beats with a simple pooled estimate of variance. This procedure is improved upon by the analysis outlined here.

The results given here differ from those reported by Kelling et al (1974). The proactive marginal mean is almost reversed from the earlier findings, while the results for control and reactive do not differ. This may be due to a different method of deseasonalizing, although personal communication indicates the methods were the same. One other possible

Table 3: Analysis of Kansas City Preventive Patrol Experiment; Reported Residence Burglaries Between October 1, 1968, and September 30, 1973 as Dependent Variable, Intervention Occurred October 1, 1972.

| Treatment | Beats 1 | 2 | 3 | 4 | 5 | $\delta_{i.}$ |
|---|---|---|---|---|---|---|
| Reactive | $\delta$=.300 | 1.709 | .517 | 1.254 | .088 | .054 |
| | $\hat{\sigma}_{\delta}$=.877 | 1.790 | 2.137 | 1.1339 | 1.298 | |
| Proactive | .315 | .696 | -.373 | 1.512 | .090 | .448 |
| | 1.018 | 1.793 | 2.017 | 1.763 | .987 | |
| Control | -.361 | 1.634 | 4.046 | 2.304 | 1.295 | 1.784 |
| | 1.423 | 1.351 | 2.162 | 1.214 | 1.329 | |
| Marginals $\bar{\delta}_{.j}$ = | -.115 | 1.346 | 1.397 | 1.690 | .491 | Grand Mean $\delta_{..}$=.962 |

df = 58 per cell

ANOVA TABLE

| Source | SS | df | MS | F | Prob. |
|---|---|---|---|---|---|
| Total | 20.157 | 885 | - | - | - |
| Treatment (T) | 5.174 | 2 | 2.587 | 2.51 | <.10 |
| Beat(B) | 6.745 | 4 | 1.686 | .74 | >.5 |
| TxB | 8.238 | 8 | 1.0298 | .45 | >.9 |
| error | - | 870 | 2.266 | - | |

explanation is that, in the current study, estimates of $\delta_{ij}$ were made using a coarser intefval on the iteration of $\theta_{ij}$ and different values of $\delta_{ij}$ could be produced. The primary purpose of this example is heuristic and these results are not definitive.

## Summary

A procedure for a two-way classification analysis of variance for time-series has been outlined. The procedure is generalizable to an N-way classification in exactly the same manner. The situation of having a large number of time-series observations must be examined and made manageable by this analysis.

## References


Bartlett, M.S. On the theoretical specification of sampling properties of autocorrelated time-series. Journal of The Royal Statistical Society (Series B), 1946, 8, 27-36.

Box, G.E.P., and Jenkins, G.M. Time-series Analysis: Forecasting and Control. San Francisco: Holden-Day, 1970.

Box, G.E.P., and Tiao, G.C. A change in the level of a non-stationary time-series. Biometrika, 52, 1965, 181-192.

Campbell, D.T., and Stanley, J.C. Experimental and Quasi-experimental Designs for Research. Chicago: Rand McNally & Co., 1966.

Durbin, J., and Watson, G.S. Testing for serial correlation in least-squares regression. I. Biometrika, 37, 1950, 409-428. II. Biometrika, 38, 1951, 159-177.

George, J.D. The effect of non-normal distributions on the integrated moving average model of time-series analysis. Unpublished, Ph.D. thesis, University of Minnesota, 1974.

Glass, G.V. Analysis of data on the Connecticut speeding crackdown as a time-series quasi-experiment. Law and Society Review, 1968, 3, 55-76.

Glass, G.V., Willson, V.L., and Gottman, J.M. The Design and Analysis of Time-Series Experiments, Boulder: Colorado Associated Universities Press, 1975.

Gullickson, A.R., Nelson, L.R., and Glass, G.V. An analysis of the violation of homogeneity of variance assumption in the IMA time-series analysis model. Paper presented at AERA Annual Meeting, February 5, 1971, New York, New York.

Kelling, G.L., Pate, T., Dieckman, D., Brown, C.E. The Kansas City Preventive Patrol Experiment, Kansas City Police Foundation, 1974 (mimeo).

Kepka, E.J. Model representation and the threat of instability in the interrupted time-series quasi-experiment. Ph.D. dissertation, Northwestern University, June 1972.

Scheffe, H. The Analysis of Variance, New York: John Wiley and Sons, Inc., 1959.

Shumway, R.H. Applied regression and analysis of variance for stationary time-series. Journal of the American Statistical Association, 1970, 65, 1527-1546.